Exhibit B

DONALD VOET
University of Pennsylvania

JUDITH G. VOET
Swarthmore College

BIOCHEMISTRY

SECOND EDITION

JOHN WILEY & SONS, INC.
New York Chichester Brisbane
Toronto Singapore

Acquisitions Editor: *Nedah Rose*
Marketing Manager: *Catherine Faduska*
Designer: *Madelyn Lesure*
Manufacturing Manager: *Susan Stetzer*
Photo Researcher: *Hilary Newman*
Illustration Coordinator: *Edward Starr*
Illustrators: *J/B Woolsey Associates*
Production Management: *Pamela Kennedy Oborski and Suzanne Ingrao*
Cover and Part Opening Illustrations © *Irving Geis*
This book was set in 9.5/12 Times Roman by Progressive Information Technologies and printed and bound by Von Hoffman Press. The cover was printed by Lehigh Press Lithographers. Color separations by Lehigh Press Colortronics.

Cover Art: Two paintings of horse heart cytochrome *c* by Irving Geis in which the protein is illuminated by its single iron atom. On the front cover the hydrophilic side chains are drawn in green, and on the back cover the hydrophobic side chains are drawn in orange. The paintings are based on an X-ray structure by Richard Dickerson.

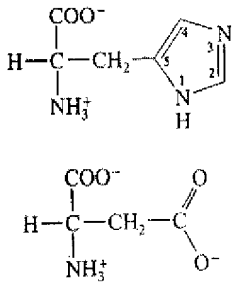
Recognizing the importance of preserving what has been written, it is a policy of John Wiley & Sons, Inc. to have books of enduring value published in the United States printed on acid-free paper, and we exert our best efforts to that end.

ISBN: 0-471-58651-X

Printed in the United States of America

10 9 8 7 6 5 4 3

CHAPTER



4

Amino Acids

It is hardly surprising that much of the early biochemical research was concerned with the study of proteins. Proteins form the class of biological macromolecules that have the most well-defined physicochemical properties and consequently they were generally easier to isolate and characterize than nucleic acids, polysaccharides, or lipids. Furthermore, proteins, particularly in the form of enzymes, have obvious biochemical functions. The central role that proteins play in biological processes has therefore been recognized since the earliest days of biochemistry. In contrast, the task of nucleic acids in the transmission and expression of genetic information was not realized until the late 1940s, the role of lipids in biological membranes was not appreciated until the 1960s, and the biological functions of polysaccharides are still somewhat mysterious.

In this chapter we study the properties of the monomeric units of proteins, the **amino acids.** It is from these substances that proteins are synthesized through processes that are discussed in Chapter 30. Amino acids are also energy metabolites and many of them are essential nutrients (Chapter 24). In addition, as we shall see, many amino acids and their derivatives are of biochemical importance in their own right (Section 4-3B).

1. THE AMINO ACIDS OF PROTEINS

The analyses of a vast number of proteins from almost every conceivable source have shown that *all proteins are composed of the 20 "standard" amino acids listed in Table 4-1.* These substances are known as **α-amino acids** because, with the exception of **proline,** they have a primary amino group and a carboxylic acid group substituent on the same carbon atom (Fig. 4-1; proline has a secondary amino group.)

A. General Properties

The pK values of the 20 "standard" α-amino acids of proteins are tabulated in Table 4-1. Here pK_1 and pK_2, respectively, refer to the α-carboxylic acid and α-amino groups, and pK_R refers to the side groups with acid–base properties. Table 4-1 indicates that the pK values of the α-carboxylic acid groups lie in a small range around 2.2 so that above pH 3.5 these groups are almost entirely in their carboxylate forms. The α-amino groups all have pK values near 9.4 and

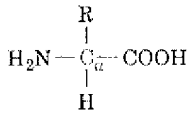


FIGURE 4-1. The general structural formula for α-amino acids. There are 20 different R groups in the commonly occurring amino acids (Table 4-1).

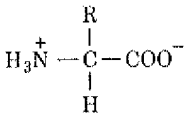


FIGURE 4-2. The zwitterionic form of the α-amino acids that occurs at physiological pH values.

are therefore almost entirely in their ammonium ion forms below pH 8.0. This leads to an important structural point: *In the physiological pH range, both the carboxylic acid and the amino groups of α-amino acids are completely ionized (Fig. 4-2).* An amino acid can therefore act as either an acid or a base. Substances with this property are said to be **amphoteric** and are referred to as **ampholytes** (*ampho*teric electro*lytes*). In Section 4-1D, we shall delve a bit deeper into the acid–base properties of the amino acids.

Molecules that bear charged groups of opposite polarity are known as **zwitterions** or **dipolar ions.** The zwitterionic character of the α-amino acids has been established by several methods including spectroscopic measurements and X-ray crystal structure determinations (in the solid state the α-amino acids are zwitterionic because the basic amine group abstracts a proton from the nearby acidic carboxylic acid group). Because amino acids are zwitterions, their physical properties are characteristic of ionic compounds. For instance, most α-amino acids have melting points near 300°C, whereas their nonionic derivatives usually melt around 100°C. Furthermore, amino acids, like other ionic compounds, are more soluble in polar solvents than in nonpolar solvents. Indeed, most α-amino acids are very soluble in water but are largely insoluble in most organic solvents.

B. Peptide Bonds

The α-amino acids polymerize, at least conceptually, through the elimination of a water molecule as is indicated in Fig. 4-3. The resulting CO—NH linkage is known as a **peptide bond.** Polymers composed of two, three, a few (3–10), and many **amino acid residues** (alternatively called **peptide units**) are known, respectively, as **dipeptides, tripeptides, oligopeptides,** and **polypeptides.** These substances, however, are often referred to simply as "peptides."

Proteins are molecules that consist of one or more polypeptide chains. These polypeptides range in length from ~40 to over 4000 amino acid residues and, since the average mass of an amino acid residue is ~110 D, have molecular masses that range from ~4 to over 440 kD.

*Polypeptides are **linear polymers;*** that is, each amino acid residue is linked to its neighbors in a head-to-tail fashion rather than forming branched chains. This observation reflects the underlying elegant simplicity of the way living systems construct these macromolecules for, as we shall see, the nucleic acids that encode the amino acid sequences of polypeptides are also linear polymers. This permits the direct correspondence between the monomer (nucleotide) sequence of a nucleic acid and the monomer (amino acid) sequence of the corresponding polypeptide without the added complication of specifying the positions and sequences of any branching chains.

With 20 different choices available for each amino acid residue in a polypeptide chain, it is easy to see that a huge number of different protein molecules can exist. For example, for dipeptides, each of the 20 different choices for the first amino acid residue can have 20 different choices for the second amino acid residue, for a total of $20^2 = 400$ distinct dipeptides. Similarly, for tripeptides, there are 20 possibilities for each of the 400 choices of dipeptides to yield a total of $20^3 = 8000$ different tripeptides. A relatively small protein molecule consists of a single polypeptide chain of 100 residues. There are $20^{100} = 1.27 \times 10^{130}$ possible unique polypeptide chains of this length, a quantity vastly greater than the estimated number of atoms in the universe (9×10^{78}). Clearly, nature can have made only a tiny fraction of the possible different protein molecules. Nevertheless, *the various organisms on Earth collectively synthesize an enormous number of different protein molecules whose great range of physicochemical characteristics stem largely from the varied properties of the 20 "standard" amino acids.*

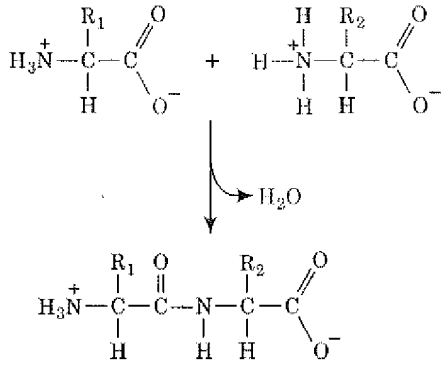
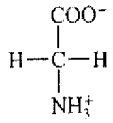
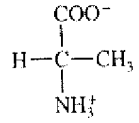
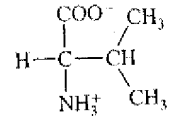
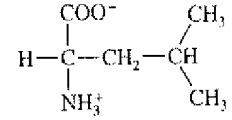
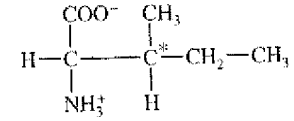
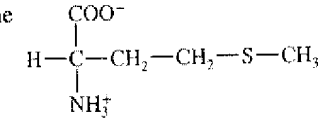
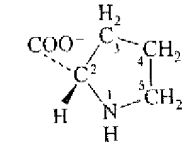
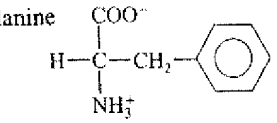
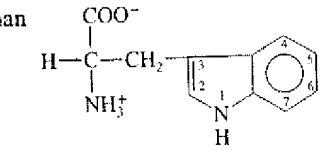


FIGURE 4-3. The condensation of two α-amino acids to form a dipeptide. The peptide bond is shown in red.

TABLE 4-1. COVALENT STRUCTURES AND ABBREVIATIONS OF THE "STANDARD" AMINO ACIDS OF PROTEINS, THEIR OCCURANCE, AND THE p*K* VALUES OF THEIR IONIZING GROUPS

Name, Three-letter Symbol, and One-letter Symbol	Structural Formula[a]	Residue Mass (D)[b]	Average Occurence in Proteins (%)[c]	pK_1 α-COOH[d]	pK_2 α-NH_3^+ [d]	pK_R Side chain[d]
Amino acids with nonpolar side chains						
Glycine Gly G		57.0	7.2	2.35	9.78	
Alanine Ala A		71.1	7.8	2.35	9.87	
Valine Val V		99.1	6.6	2.29	9.74	
Leucine Leu L		113.2	9.1	2.33	9.74	
Isoleucine Ile I		113.2	5.3	2.32	9.76	
Methionine Met M		131.2	2.2	2.13	9.28	
Proline Pro P		97.1	5.2	1.95	10.64	
Phenylalanine Phe F		147.2	3.9	2.20	9.31	
Tryptophan Trp W		186.2	1.4	2.46	9.41	

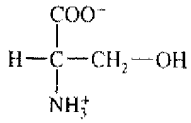
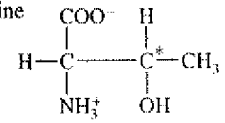
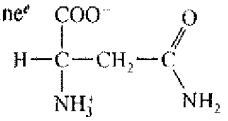
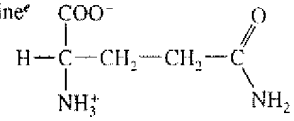
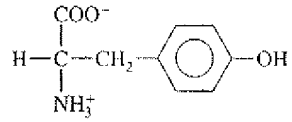
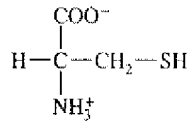
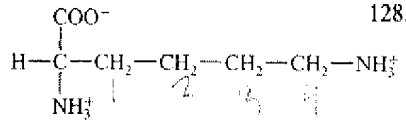
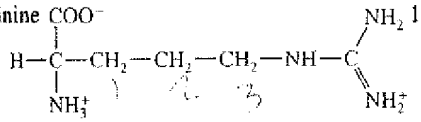
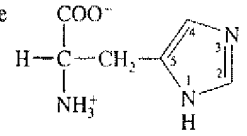
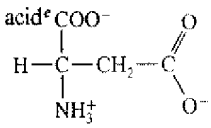
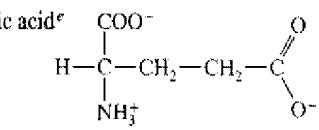
[a] The ionic forms shown are those predominating at pH 7.0 although residue mass is given for the neutral compound. The C_α atoms, as well as those atoms marked with an asterisk, are chiral centers with configurations as indicated according to Fischer projection formulas. The standard organic numbering system is provided for heterocycles.

[b] The residue masses are given for the neutral residues. For the molecular masses of the parent amino acids, add 18.0 D, the molecular mass of H_2O, to the residue masses. For side chain masses, subtract 56.0 D, the formula mass of a peptide group, from the residue masses.

[c] Calculated from a database of nonredundant proteins containing 300,688 residues as compiled by Doolittle, R. F. *in* Fasman, G. D. (Ed.), *Predictions of Protein Structure and the Principles of Protein Conformation*, Plenum Press (1989).

[d] *Source:* Dawson, R.M.C., Elliott, D.C., Elliott, W.H. and Jones, K.M., *Data for Biochemical Research* (3rd ed.), *pp.* 1–31, Oxford Science Publications (1986).

[e] The three- and one-letter symbols for asparagine *or* aspartic acid are Asx and B, whereas for glutamine *or* glutamic acid they are Glx and Z. The one-letter symbol for an undetermined or "nonstandard" amino acid is X.

Name, Three-letter Symbol, and One-letter Symbol	Structural Formula[a]	Residue Mass (D)[b]	Average Occurence in Proteins (%)[c]	pK_1 α-COOH[d]	pK_2 α-NH_3^+[d]	pK_R Side chain[d]
Amino acids with uncharged polar side chains						
Serine Ser S		87.1	6.8	2.19	9.21	
Threonine Thr T		101.1	5.9	2.09	9.10	
Asparagine[e] Asn N		114.1	4.3	2.14	8.72	
Glutamine[e] Gln Q		128.1	4.3	2.17	9.13	
Tyrosine Tyr Y		163.2	3.2	2.20	9.21	10.46 (phenol)
Cysteine Cys C		103.1	1.9	1.92	10.70	8.37 (sulfhydryl)
Amino acids with charged polar side chains						
Lysine Lys K		128.2	5.9	2.16	9.06	10.54 (ε-NH_3^+)
Arginine Arg R		156.2	5.1	1.82	8.99	12.48 (guanidino)
Histidine His H		137.1	2.3	1.80	9.33	6.04 (imidazole)
Aspartic acid[e] Asp D		115.1	5.3	1.99	9.90	3.90 (β-COOH)
Glutamic acid[e] Glu E		129.1	6.3	2.10	9.47	4.07 (γ-COOH)